DIALS: Diffraction Integration for Advanced Light Sources. Workshop 3
LBNL, 11 February 2013

Challenges Ahead: What's on the Horizon?

Gwyndaf Evans
Diamond Light Source, UK

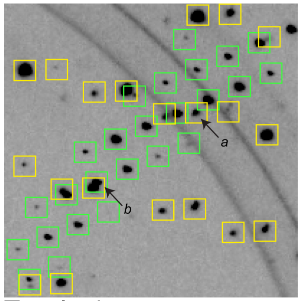
Nick Sauter
Lawrence Berkeley National Laboratory, US

Diffraction Integration for Advanced Light Sources

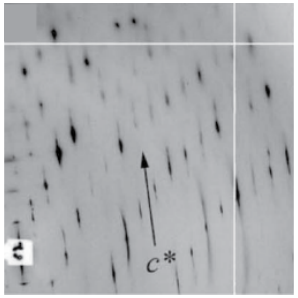
- New data reduction software to match evolving detectors & synchrotron sources.
- Separately funded projects
  - EU BioStruct-X Work Package 6
  - NIH / National Institute of General Medical Sciences
- Collaborative development & common approaches
- Open source (cctbx.sf.net; dials.sf.net) & open architecture
- Participation encouraged by beamline developers & detector manufacturers
- Contributions of code and ideas welcome
- DIALS will be a platform for research on data reduction methods. Seek to document the best practices so that algorithms become transparent.
- Developer workshops (2012: Diamond & Boston; Feb 2013: Berkeley)
- Future user workshops

Basic Deliverables & Longer Term Aims

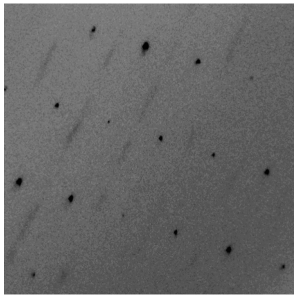
- Basic:
  - Application programming interface to Bragg diffraction data reduction: postrefinement and integration.
  - *Integration should take no longer than the time taken to measure the data.* Use of available computing resources including multiprocessing & GPU. (Peter Zwart, Jon Schuermann, Tim McPhillips)
  - Incorporate beamline-specific & detector-specific information
- Advanced:
  - Address difficult problems such as overlapping spots & multiple lattices
  - Physics-based modeling of Bragg spot shape including ray-tracing approaches (James Holton, David Waterman, James Parkhurst)



Two lattices

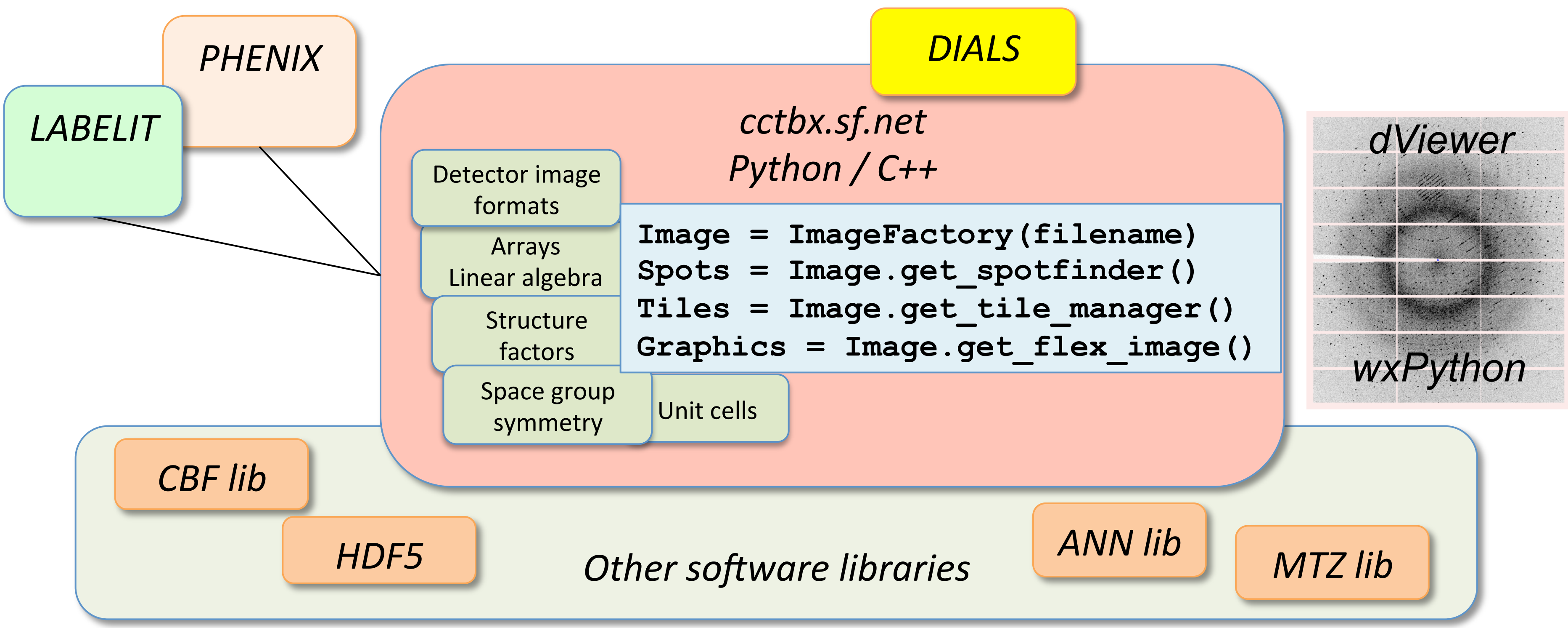


Lattice-translocation disorder
Yeates (2009) Acta D


Order-disorder twinning
Dauter (2009) Acta D

Software Architecture and Management

- Based on the LBNL Computational Crystallography Toolbox: Python / C++ (Nat Echols, Richard Gildea)
- Open source; BSD-style license allows distribution of derivative works
- Others can modify source code & contribute to the official version
- Run-time discovery of user-supplied plug-ins as done in XIA2 (Graeme Winter)
- Version control, nightly build, unit tests, regression tests, cross-platform, downloads



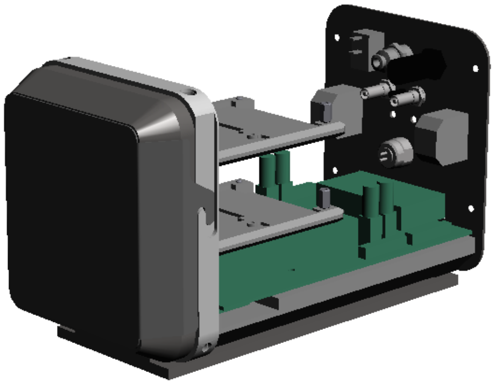
Dectris Has Licensed the EIGER Chip from PSI

Chip Features

	PILATUS3	**EIGER**
Radiation tolerance	Radiation tolerant design	Same
Chip size	17.54 x 10.45 mm^2	19.3 x 20.1 mm^2 -> 2 x
Pixel size	172 x 172 μm^2	75 x 75 μm^2 -> / 5.3
Pixel array	60 x 97 = 5820	256 x 256 = 65536 -> 11.3 x
Count rate	$10*10^6$ x-rays/pixel/s	$2*10^6$ x-rays/pixel/s
Counter	20 bit	12 bit (8 and 4 bit mode)
Continuous readout	No	Yes
Clock speed	150 MHz	100 MHz DDR
Threshold adjustment	6 bit DAC	Same
Overflow control	Yes	Yes

16M Detector to be Delivered June 2015

EIGER Product Family



The EIGER systems are developed in the following configurations:

Type	Modules	Active area [mm²]	Pixel array	Frame rate [Hz]	DCB-PC [Gbps]
1M	2	77 x 80	1030 x 1065	3000	40
4M	8	155 x 163	2070 x 2167	750	40
9M	18	233 x 245	3110 x 3269	333	40
16M	32	311 x 328	4150 x 4371	187	40

Discussion of IT: Heiner Billich, PSI

- Internal uncompressed 6.5 GB/s; external compressed 800 MB/s
- 800 MB/s stays within reach of 10Gb Ethernet.
- Need fast algorithms, parallel compression, factor 7-8.
- Single data connection, single client to few files.
- Well understood data access pattern is sequential large block IO.
- Works well with export, archive and backup
- Single file per image is completely at odds with existing solutions to seq. large block IO; latency of meta ops like create, open, close, stat.
- HDF5 supports large block IO very well & is used in HPC for efficient storate of large datasets. May 2013 release supports parallel compression external to HDF5, and dynamic load of custom filters.
- HDF5 is like a file system; Nexus is a domain-specific format for metadata

Challenges of Data Access at 10 TB/ hour

- Multiple images written to single file container.
- LCLS' XTC serial trajectory file paradigm works well for diffract-and-destroy experiments: each crystal is a separate experiment.
- Need random access for indexing of goniometer data: HDF5 seems to be a good choice for synchrotron experiments, provided it is fast enough.
- Data will be accessed for image visualization, Bragg spot statistics, data reduction, rad damage assessment. Heavy parallel access essential.
- Cannot say at this time whether on-the-fly stream API or HDF5 file I/O is preferred. Should engineer both solutions.

Acknowledgements

Diamond Light Source

Gwyndaf Evans
Graeme Winter
James Parkhurst
Luis Fuentes-Montero

Berkeley National Lab

Nick Sauter
Johan Hattne
Richard Gildea
Aaron Brewster
Paul Adams

CCP4

David Waterman

Work funded by EU BioStruct-X, Diamond, CCP4, NIH/NIGMS, DOE